

Implementación de una base de datos relacional difusa. Caso práctico: tutoría académica

(Implementation of a fuzzy relational database. Case study: academic tutoring)

Ciro Saguay¹, Rodrigo Proaño², Bolívar Jácome³, Denisse Aguirre⁴

Resumen:

El presente artículo describe el proceso de implementación de la base de datos relacional difusa para el caso práctico de las tutorías académicas de la Facultad de Ciencias de la Ingeniería de la Universidad Tecnológica Equinoccial (UTE). Para la implementación se utilizó como metodología la arquitectura ANSI-SPARC de bases de datos, la misma que abstrae la información en niveles, en el nivel externo se obtuvo los requerimientos funcionales, en el nivel conceptual se obtuvo el modelo relacional difuso, para lograr este modelo se realizó la transformación de los datos difusos a través de modelos matemáticos usando la herramienta Fuzzy-Lookup y en el nivel físico se implementó la base de datos relacional difusa, además se desarrolló una interfaz en lenguaje Java por medio de la cual se ingresan datos y se realizan consultas a la base de datos relacional difusa para verificar su funcionamiento.

Palabras clave: Base de datos relacional difusa; Implementación de bases de datos relacional difusa; lógica difusa.

Abstract:

This paper describes the process of implementation of a diffused relational database in the practical case of the academic tutorials of the Faculty of Engineering Sciences of the Equinoctial Technological University (UTE). For the implementation, the ANSI-SPARC database architecture was used as the methodology, which abstracts the information into levels, at the external level the functional requirements were obtained, at the conceptual level, the diffused relational model was obtained. To achieve this model, we performed the transformation of the diffuse data through mathematical models using the Fuzzy-Lookup tool and at the physical level the diffused relational database was implemented. In addition, an user interface was developed using Java through which data is entered and queries are made to the diffused relational database to verify its operation.

Keywords: Fuzzy relational data bases; Implementation of fuzzy relational data bases; fuzzy logic.

¹ Universidad Tecnológica Equinoccial, Quito-Ecuador (csaguay@ute.edu.ec)

² Universidad Tecnológica Equinoccial, Quito-Ecuador (rodrigo.proano@ute.edu.ec)

³ Universidad Tecnológica Equinoccial, Quito-Ecuador (sbolivar.jacome@ute.edu.ec)

⁴ Universidad Tecnológica Equinoccial, Quito-Ecuador (carolina_aguirre87@outlook.com)

1. Introducción

En bases de datos, la lógica difusa es utilizada para extender las bases de datos relacionales (BDR) a difusas (BDRD) (Barranco, C.D., Campaña, J.R., & Medina, J.M., 2008), estas bases permiten el almacenamiento y procesamiento de datos clásicos y difusos de manera conjunta, para lo cual los datos difusos deben ser sujetos de un proceso previo de traducción o transformación para asumir grados de pertenencia en base a funciones lineales o regresión lineal, y luego ser implementados con herramientas de procesamiento de datos, estas bases de datos brindan flexibilidad a aquellas empresas o instituciones donde se manipula información imprecisa y que no puede ser considerada de manera real para cálculos matemáticos o comparaciones estadísticas debido a la escasa información numérica que proveen.

En la teoría de lógica difusa, propuesta por Zadeh en 1965 se menciona el concepto de conjuntos difusos para explicar el sentido de pertenencia de los elementos que forman parte de este tipo de conjuntos. Un conjunto difuso A es parte o subconjunto de un conjunto difuso B si cumple con la siguiente condición: $\mu_A(x) \leq \mu_B(x) \quad \forall x \in X$. (Pérez & León, 2007)

Para la definición de pertenencia, se establecen reglas que serán consideradas como indicadores de referencia de lo que un elemento debe cumplir para pertenecer a un concepto dado, de esta forma, es posible guiarse a través de enunciados en lenguaje natural para conocer el grado de pertenencia de cada elemento de un conjunto difuso, también se debe mencionar que, para operar posteriormente con estos elementos difusos, existen etiquetas las cuales se pueden reconocer por su tipo. (Pérez & León, 2007)

SCIPEDIA

Register for free at <https://www.scipedia.com> to download the version without the watermark

La tutoría académica de la UTE se la realiza a través de entrevistas, en la misma constan preguntas que buscan conocer detalles sobre los niveles de concentración de los estudiantes, el tiempo que consideran para la preparación de sus exámenes, la relación que tienen con los docentes de las asignaturas en conflicto, así como también temas personales de relevancia que pudieran ser causas de un mal desempeño académico (G. Bravo, 2012), obteniéndose atributos y valores imprecisos e inciertos en su mayoría, por lo que no se pueden plasmar de forma veraz y adecuada en una base de datos relacional.

El artículo describe la implementación de la base de datos relacional difusa de la tutoría académica de la Facultad de Ciencias de Ingeniería e Industrias de la UTE utilizando etiquetas tipo I, a través de la herramienta Fuzzy Lookup (Network, 2016) que es parte de los servicios de integración de SQL Server (SQL Server Integration Services – SSIS), permitiendo asignar un valor de pertenencia a aquellos datos difusos en base a su naturaleza que sería considerado como el

conjunto universo. Finalmente, se desarrolla una interfaz que permite recolectar datos clásicos y difusos y permite obtener resultados que respondan a demandas difusas específicas.

2. Metodología

Para la implementación de la base de datos relacional difusa se utilizó la arquitectura ANSI/X3/SPARC (American National Standard Institute - Standards Planning And Requirements Committee), que utiliza modelos de datos en los diferentes niveles para lograr independencia (A. Silberschatz, 2002), además en cada nivel se incorporan elementos propios de las bases de datos difusas. A continuación se explica lo realizado en cada nivel.

Nivel externo.- en esta etapa a través de un estudio exploratorio y descriptivo se realizó el análisis de requisitos de información que forma parte de las tutorías académicas y que será almacenada en la base de datos relacional difusa, como resultado de esto se identificó información clásica y difusa.

Nivel conceptual.- en esta etapa se maneja la estructura conceptual de la base de datos, se identifican los elementos lógicos, tales como las entidades, sus atributos y las relaciones entre dichas entidades, además se consideró la relación existente entre datos clásicos y difusos en los tres niveles de modelamiento de datos difusos, es decir, en la lógica del negocio o la empresa, en cuanto a los datos por sí mismos como sujetos de procesamiento y en el lenguaje de consultas, para determinar si son procesados de forma tradicional o si deben pasar previamente por un proceso de traducción, que es el caso de los datos difusos.

Register for free at <https://www.scipedia.com> to download the version without the watermark

Una vez definido la estructura de los datos y clasificarlos en clásicos y difusos, se procedió a establecer el modelo difuso establecido por GEFRED (R.M. Cicília, 2011), el cual incluye las siguientes etapas:

Etapas de Fusificación.- En esta etapa se definió las funciones de relevancia por medio de las cuales se busca establecer los grados de pertenencia a los elementos del conjunto universo.

Reglas Base.- Se establecieron las reglas por las cuales se realizaron las traducciones o transformaciones de los datos difusos para ser procesados con los datos clásicos.

Etapas de Inferencia.- Se evaluó cada uno de los conjuntos difusos y se combinó con la información definida dentro de las reglas base.

Etapas de Difusificación.- Etapa final en la cual se obtuvo los resultados de la traducción y procesamiento de los datos difusos.

Estas etapas implican el uso de métodos matemáticos para establecer los conjuntos difusos y los grados de pertenencia que corresponden a las etiquetas lingüísticas que son elementos de dichos

conjuntos, el agrupar de esta manera permite a su vez reconocer aquellas entidades difusas e integrarlas y relacionarlas según sus atributos difusos en el modelo conceptual.

Nivel interno.- una vez definido el modelo conceptual de la base de datos, se utilizó la herramienta CASE Power Designer para generar el script SQL para la implementación en el DBMS SQL Server y para la transformación de las entidades difusas se utilizó la herramienta Fuzzy Lookup del mismo DBMS, esta herramienta requiere de tablas de referencias para realizar este proceso, estas tablas se agregaron en el modelo conceptual sirviendo de catálogo de búsqueda para encontrar coincidencias entre los datos difusos que ingresan al proceso de transformación y los datos de referencia que serán ingresados en base a los grados de pertenencia calculados en el nivel conceptual.

La herramienta Fuzzy Lookup genera como resultado dos tipos de puntuaciones por cada búsqueda: Puntuación de Similitud y Puntuación de Confianza, en este caso se definió el modelo de unificación mediante relaciones de similitud para manejar de forma precisa tanto la información de entrada, es decir las demandas, como los resultados después de haber pasado por el proceso de transformación y comparación de la puntuación de similitud obtenido entre el dato difuso y las tablas de referencias.

Estas etapas implicaron el uso de la técnica de regresión lineal con el método de mínimos cuadrados para establecer los conjuntos difusos y los grados de pertenencia. (D. Anderson, 2008)

El método de mínimos cuadrados es empleado para encontrar la ecuación de regresión estimada a través de datos tomados de una muestra.

Register for free at <https://www.scipedia.com> to download the version without the watermark

Al graficar un conjunto de datos de una muestra, se obtiene un gráfico de dispersión, por lo que se aplica un método de regresión lineal para encontrar la recta que se ajuste a los datos de la muestra.

La ecuación de regresión lineal simple está dada en la siguiente expresión matemática:

$$\hat{y}_i = b_0 + b_1 x_i$$

Dónde:

\hat{y}_i = Valor estimado de grados de pertenencia.

b_0 = Intersección de recta de regresión con el eje y.

b_1 = Pendiente de la recta de regresión.

x_i = Definición de atributos difusos.

La definición de cada uno de los elementos de la expresión antes descrita, está dada en base al caso particular del análisis de datos difusos.

El criterio del método de mínimos cuadrados, se muestra en la siguiente expresión:

$$\min \Sigma (y_i - \hat{y}_i)^2$$

Dónde:

y_i = Valor real de la variable dependiente.

\hat{y}_i = Valor estimado de la variable independiente.

Al sustituir la ecuación de regresión lineal simple en la expresión matemática de mínimos cuadrados, se obtiene la siguiente expresión:

$$\Sigma (y_i - b_0 - b_1 x_i)^2$$

Para el desarrollo de la interfaz se basará en el ciclo de vida del modelo espiral propuesto por (Pressaman, 2010), este modelo describe el desarrollo en un corto plazo, donde el usuario final y sus requisitos se encuentran definidos, para de esta manera tener lo más pronto posible un ejecutable que pueda ser probado por el usuario final. Este modelo está compuesto por las siguientes fases:

- Análisis de interfaz
- Diseño de interfaz
- Desarrollo de interfaz
- Validación de interfaz

SCIPEDIA

3. Resultados

Register for free at <https://www.scipedia.com> to download the version without the watermark

Como parte del estudio exploratorio y descriptivo de la tutoría académica, se revisó el manual interno de tutorías académicas (SITUTOR) (G. Bravo, 2012), se determinó que el proceso se inicia con la planificación de la entrevista de los estudiantes de una carrera, la información se registra en una hoja de Excel parametrizada, esta se encuentra formada por 21 preguntas referentes a la información personal del estudiante e información general sobre la carrera y las asignaturas en las cuales registra segunda y tercera matrícula. A través de 16 preguntas, se analizan aspectos que afectan el desempeño académico del estudiante, se toma en cuenta hábitos de estudio, ambiente de estudio, actividades adicionales y la relación que existe con los docentes, entendiendo por relación al estado de confianza, respeto y comprensión entre las partes durante el tiempo de clases. El análisis de conocimientos se basa en 4 preguntas fundamentales: conocimientos de informática, desenvolvimiento en internet, manejo de la plataforma o campus virtual de la UTE y conocimiento del idioma Inglés. El estudiante a través del dialogo con el tutor proporciona aspectos como nivel de comunicación, grado de responsabilidad, capacidad de trabajar en equipo y una última pregunta que permite conocer cuáles son las expectativas del estudiante con respecto a las tutorías académicas.

Del análisis realizado se identificó datos clásicos y difusos de las tutorías académicas, información parcial de estos se muestran en la *Tabla 1*.

Tabla 1. Clasificación de datos clásicos y difusos (parcial)

Pregunta	Dato	Tipo de dato
¿Qué espera el estudiante de la tutoría?	Apoyo Profesores	Clásico
	Guía en trámites	
	Apoyo en las notas	
	Apoyo Personal	
	Apoyo Académico	
Relaciones Familiares	Excelente	Difuso
	Buena	
	Mala	
Relaciones Compañeros	Excelente	Difuso
	Buena	
	Mala	
Relaciones Pareja	Excelente	Difuso
	Buena	
	Mala	
	No tiene	Clásico

El universo de datos de la tutoría académica permite clasificar y englobar los datos difusos en tres conjuntos:

- Conjunto Difuso 1: Alto, medio, bajo.
- Conjunto Difuso 2: Excelente, buena, mala.
- Conjunto Difuso 3: Nada, poco, bastante, mucho.

Register for free at <https://www.scipedia.com> to download the version without the watermark

Estos tres conjuntos de datos difusos, son procesados para obtener el modelo difuso, como ejemplo se muestra una parte del proceso en cada etapa.

Etapas de Fusificación.- la referencia del grado de pertenencia dentro del universo de datos difusos, en primera instancia se lo realizó en lenguaje natural ejemplo:

- Niveles de concentración (alto, medio, bajo)
- Relación con el docente (excelente, buena, mala)
- Análisis de conocimientos – informática, internet, plataforma, idioma inglés (nada, poco, bastante, mucho)

Reglas Base.- Cada conjunto de datos difusos fue tratado por separado para la transformación matemática, debido a que representan una escala diferente de grados de pertenencia, los cuales están dentro de un rango de $[0,1]$. En el caso práctico de las tutorías académicas, los conjuntos de datos difusos se encuentra definidos con un número de 3 a 4 datos por conjunto, para definir con

mayor claridad los grados de pertenencia, se procede a colocar atributos difusos intermedios entre los atributos, estos datos intermedios son representados básicamente como puntos de referencia dentro de las gráficas de dispersión y ajuste de recta bajo el criterio de mínimos cuadrados.

Etapa de Inferencia.- Dentro de las reglas base se declaró tres reglas fundamentales, que se aplican a los conjuntos de datos difusos, para tener un mejor entendimiento se utilizan gráficos de dispersión, ubicando en el eje horizontal (x) los atributos difusos, y en el eje vertical (y) los grados de pertenencia, las reglas aplicadas son:

- Regla Base 1: Transformación matemática por separado de cada conjunto de atributos difusos.
- Regla Base 2: Los grados de pertenencia a definirse deben estar dentro del rango [0,1].
- Regla Base 3: Colocar atributos difusos intermedios como puntos de referencia.

Como ejemplo se muestra la aplicación de estas reglas de inferencia al conjunto difuso 1 (Alto, medio, bajo) en la *Figura 1*.



Figura 1. Inferencia Conjunto Difuso 1

Etapa de Difusificación.- Una vez estimados los grados de pertenencia de cada conjunto de atributos difusos, se inicia el procesamiento de dichos valores a través de la regresión lineal simple con el método de mínimos cuadrados. Como ejemplo los valores obtenidos para el Conjunto Difuso 1 (Alto, medio, bajo) se muestran en la *Tabla 2*.

Tabla 2. Cálculo de pendiente y punto de intersección de la recta, Conjunto Difuso 1

Núm. Observación	x_i	y_i	$x_i - \bar{x}$	$y_i - \bar{y}$	$(x_i - \bar{x})(y_i - \bar{y})$	$(x_i - \bar{x})^2$
1	0,2	0,1	-0,32	-0,4	0,128	0,1024
2	0,4	0,3	-0,12	-0,2	0,024	0,0144
3	0,5	0,4	-0,02	-0,1	0,002	0,0004
4	0,7	0,8	0,18	0,3	0,054	0,0324
5	0,8	0,9	0,28	0,4	0,112	0,0784
Totales	2,6	2,5			0,32	0,228
	Σx_i	Σy_i			$\Sigma (x_i - \bar{x})(y_i - \bar{y})$	$\Sigma (x_i - \bar{x})^2$

La gráfica de la recta ajustada a los puntos del Conjunto Difuso 1 se muestra en la *Figura 2*.

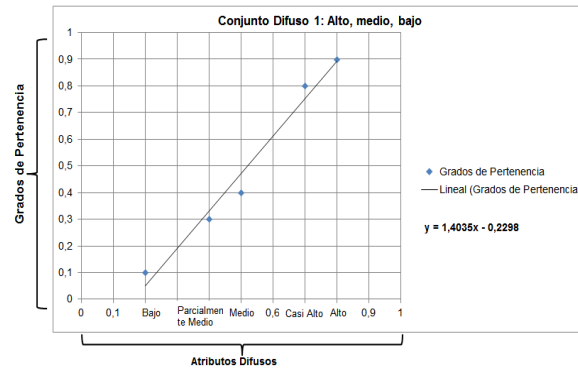


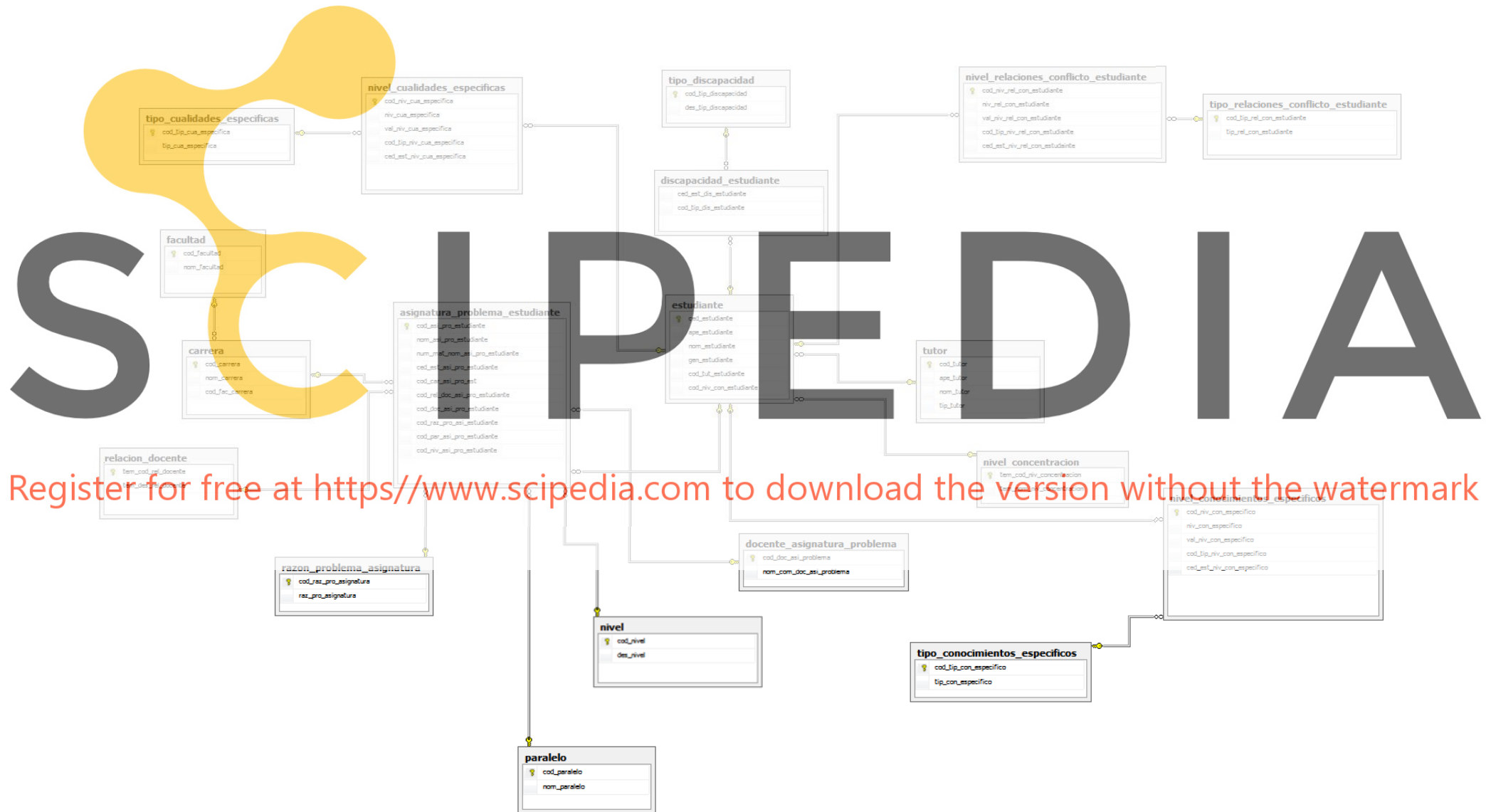
Figura 2. Recta ajustada del Conjunto Difuso 1

El modelo relacional difuso consta de 19 tablas que contienen atributos clásicos (Crisp) y atributos difusos como se indica en la *Tabla 3*.

Tabla 3. Tablas que intervienen en el modelo relacional difuso

Nombre de Tabla	Tipo atributo
Facultad	Crisp
Carrera	Crisp
Nivel	Crisp
Paralelo	Crisp
asignatura_problema_estudiante	Crisp
docente_asignatura_problema	Crisp
relacion_docente	Crisp / Difusos
razon_problema_asignatura	Crisp
tipo_discapacidad	Crisp
discapacidad_estudiante	Crisp
Estudiante	Crisp
tipo_relaciones_conflicto_estudiante	Crisp
nivel_relaciones_conflicto_estudiante	Crisp / Difusos
tutor	Crisp
nivel_concentracion	Crisp / Difusos
tipo_cualidad_especifica	Crisp
nivel_cualidades_especificas	Crisp / Difusos
tipo_conocimiento_especifico	Crisp
nivel_conocimiento_especifico	Crisp / Difusos

Las tablas que contienen atributos clásicos (Crisp) son tratados dentro de un modelo Entidad – Relación común, mientras que las tablas que contienen atributos difusos, son sujetos de transformación a través de etiquetas lingüísticas almacenadas en lenguaje natural dentro de las respectivas tablas. El esquema final de la base de datos se presenta en la figura 3, el mismo que consta de 19 tablas con datos clásicos y difusos.



Register for free at <https://www.scipedia.com> to download the version without the watermark

Figura 3. Modelo de base de datos

El proceso de transformación se realizó mediante la herramienta Fuzzy Lookup que es parte de los servicios de integración de SQL Server (SQL Server Integration Services – SSIS), este proceso requiere de una tabla de origen que contiene la información difusa recolectada a través de las entrevistas y la tabla de referencia que contiene los atributos difusos con sus grados de pertenencia encontrados en el proceso de transformación matemático.

Las tablas de referencia de aquellas tablas que contienen atributos difusos, se muestran en la *Tabla 4*.

Tabla 4. Tablas de referencia del modelo relacional difuso

Nombre de Tabla	Tipo de Atributos
referencia_nivel_concentracion	Crisp / Difusos
referencia_relacion_docente	Crisp / Difusos
referencia_conocimientos_especificos	Crisp / Difusos
referencia_cualidades_especificas	Crisp / Difusos
referencia_relaciones_conflicto_estudiante	Crisp / Difusos

A continuación se muestra un ejemplo SQL para la creación de la tabla *nivel de concentración*, la misma que va a contener datos clásicos y difusos:

```
CREATE TABLE nivel_concentracion (
cod_niv_concentracion INT IDENTITY (1,1) not null PRIMARY KEY,
nom_niv_concentracion VARCHAR (50) not null,
val_niv_concentracion DECIMAL (2,2))
```

Register for free at <http://www.scipedia.com> to download the version without the watermark

aquellas tablas difusas, cuentan con un campo donde se registrará el valor numérico resultante de la transformación difusa de las etiquetas lingüísticas registradas en lenguaje natural, de manera provisional, este campo se registra inicialmente con el valor de cero (0.00).

The screenshot shows the SQL Server Enterprise Manager interface. The 'Results' pane displays the output of a query that selects data from several tables: nivel_concentracion, relacion_docente, conocimientos_especificos, cualidades_especificas, and relaciones_conflicto_estudiante. The tables contain linguistic labels and their corresponding numerical values (0.00).

cod_niv_concentracion	nom_niv_concentracion	val_niv_concentracion
1	MALA	0.00
2	RELATIVAMENTE BUENA	0.00
3	BUENA	0.00
4	PRACTICAMENTE EXCELENTE	0.00

cod_rel_docente	nom_rel_docente	val_rel_docente
1	MALA	0.00
2	RELATIVAMENTE BUENA	0.00
3	BUENA	0.00
4	PRACTICAMENTE EXCELENTE	0.00

cod_niv_con_especifico	nom_niv_con_especifico	val_niv_con_especifico
1	BASTANTE	0.00
2	NADA	0.00
3	CASI NADA	0.00
4	MUCHO	0.00
5	CASI NADA	0.00

cod_niv_cua_especifica	nom_niv_cua_especifica	val_niv_cua_especifica
1	PARCIALMENTE MEDIO	0.00
2	CASI ALTO	0.00
3	ALTO	0.00

cod_niv_rel_estudiante	nom_niv_rel_estudiante	val_niv_rel_estudiante
1	RELATIVAMENTE BUENA	0.00
2	PRACTICAMENTE EXCELENTE	0.00
3	PRACTICAMENTE EXCELENTE	0.00
4	RELATIVAMENTE BUENA	0.00
5	EXCELENTE	0.00
6	BUENA	0.00
7	RELATIVAMENTE BUENA	0.00
8	MALA	0.00

Figura 4. Estado inicial-tablas difusas

Posterior a la creación de las tablas se procedió a ingresar la información. Un ejemplo de inserción se muestra para la tabla nivel de concentración.

```
insert into referencia_nivel_concentracion values ('BAJO', 0.05)
insert into referencia_nivel_concentracion values ('MEDIO', 0.47)
insert into referencia_nivel_concentracion values ('ALTO', 0.90)
```

El flujo de transformación difusa para cada tabla que contiene atributos difusos se refleja en la *Figura 5*, las tablas de origen, pasan por una transformación difusa, derivación de columnas y ejecución de comandos SQL para finalmente desembocar en las tablas de destino que son las mismas tablas, pero internamente modificadas por la transformación.

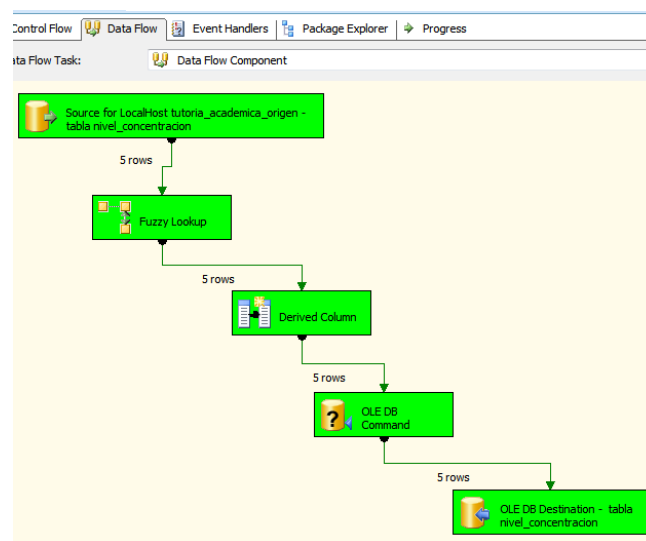


Figura 5. Flujo de transformación difusa

Posterior a la ejecución del flujo de transformación se comprueba que, los registros difusos de la base de datos ya cuentan con los valores numéricos equivalentes a las etiquetas lingüísticas, los registros que han sido transformados pueden ahora ser consultados y visualizados, con la sentencia *select* respectiva del lenguaje SQL como se muestra en la *Figura 6*.

SQL Query1.sql - LINA AGUIRRE (551)*

```

select * from nivel_concentracion
select * from relacion_docente
select * from nivel_conocimientos_especificos
select * from nivel_qualidades_especificas
select * from nivel_relaciones_conflicto_estudiante

```

cod_niv_concentracion		nom_niv_concentracion	val_niv_concentracion
1	1	BAJO	0.05
2	2	PARCIALMENTE MEDIO	0.47
3	3	MEDIO	0.47
4	4	CASI ALTO	0.90
5	5	ALTO	0.90

cod_rel_docente		des_rel_docente	val_rel_docente
1	1	MALA	0.19
2	2	RELATIVAMENTE BUENA	0.56
3	3	BUENA	0.56
4	4	PRACTICAMENTE EXCELENTE	0.93
5	5	EXCELENTE	0.93

cod_niv_con_especifico		niv_con_especifico	val_niv_con_especifico	cod_ip_niv_con_especifico	ced_est_niv_con_especifico
1	1	BASTANTE	0.61	1	1712166907
2	2	NADA	0.03	2	1712166907
3	3	CASI NADA	0.03	3	1712166907
4	4	MUCHO	0.90	4	1712166907

cod_niv_cua_especifico		niv_cua_especifico	val_niv_cua_especifico	cod_ip_niv_cua_especifico	ced_est_niv_cua_especifico
1	1	PARCIALMENTE MEDIO	0.47	1	1712166907
2	2	CASI ALTO	0.90	2	1712166907
3	3	ALTO	0.90	3	1712166907
4	4	CASI ALTO	0.90	4	1712166907

cod_niv_rel_con_estudiante		niv_rel_con_estudiante	val_niv_rel_con_estudiante	cod_ip_niv_rel_con_estudiante	ced_est_niv_rel_con_estudiante
1	1	RELATIVAMENTE BUENA	0.56	1	1712166907
2	2	PRACTICAMENTE EXCELENTE	0.93	2	1712166907
3	3	PRACTICAMENTE EXCELENTE	0.93	3	1712166907
4	4	RELATIVAMENTE BUENA	0.56	1	1234567890

Figura 6. Comprobación de transformación difusa.

En otros sistemas gestores de bases de datos es necesario contar con un servidor FSQL (A. Urrutia, 2008) que permita a través de un lenguaje extendido realizar consultas difusas, un ejemplo de este lenguaje es la siguiente consulta:

```
SELECT *
FROM nivel_concentracion
WHERE nom_niv_concentracion FEQ $MEDIO THOLD 0.5;
```

Cada vez que sea necesario ejecutar la transformación difusa, se debe ejecutar un proceso de limpieza de atributos difusos, que consiste en “encerar” el campo de valores numéricos equivalentes a las etiquetas lingüísticas de las tablas a ser transformadas, con el objetivo de preservar la integridad de los datos difusos anteriores al integrar nuevos datos a las tablas difusas y transformar de manera global todos los datos que contenga las tablas, este proceso de limpieza consiste en ejecutar un procedimiento almacenado creado en la base de datos el cual se ejecuta directamente desde el SQLServer o desde SSIS mediante un flujo independiente a los flujos de transformación difusa, a continuación se muestra el procedimiento almacenado.

```
CREATE PROCEDURE limpiar_difusos
AS
BEGIN
UPDATE dbo.nivel_conocimientos_especificos SET val_niv_con_especifico =0
UPDATE dbo.nivel_cualidades_especificas SET val_niv_cua_especifica =0
UPDATE dbo.nivel_relaciones_conflicto_estudiante SET val_niv_rel_con_estudiante =0
END
```

Al aplicar el modelo de espiral para el desarrollo de la aplicación, se inició con el análisis de la interfaz en seis secciones, luego se procedió a diseñar las interfaces de las secciones, el diseño incluye un menú principal para guiar al usuario hacia la interfaz de ingreso de nuevas entrevistas o hacia la interfaz de consultas difusas. La *Figura 7* muestra la interfaz principal.



Figura 7. Diseño interfaz – Menú Principal

Para verificar el funcionamiento de la aplicación, se realizó pruebas ingresando varias entrevistas de estudiantes de diferentes carreras de la Facultad de Ciencias de la Ingeniería, en la *Figura 8* se

puede observar una de las consultas difusas ejecutadas antes del proceso de transformación, se puede constatar que los valores numéricos de equivalencias de las etiquetas lingüísticas son cero (0.00).

CEDULA	NOMBRES ESTUDI.	GENERO	CARRERA	PARALELO	SEMESTRE	ASIGNATU...	TIPO REL...	NIVEL REL...	VALOR N...	TIPO CUA...	NIVEL C...	VALOR N...
1234567890	SANCHEZ ANDRA...	MASCULI...	INGENIER...	B	9	ARQUITEC...	COMPAÑ...	EXCELENTE	0.00	TRABAJO...	CASI ALTO	0.00
1234567890	SANCHEZ ANDRA...	MASCULI...	INGENIER...	B	9	ROBOTICA	COMPAÑ...	EXCELENTE	0.00	TRABAJO...	CASI ALTO	0.00

Figura 8. Consulta mediante aplicación antes de transformación difusa

Posterior al proceso de transformación, en la *Figura 9* se puede observar que la misma consulta arroja los valores resultantes de la transformación difusa, con lo cual esta información ya puede ser procesada para comparaciones numéricas o estadísticos según lo requieran la comunidad de tutores académicos.

CEDULA	NOMBRES ESTU...	GENERO	CARRERA	PARALELO	SEMESTRE	ASIGNATU...	TIPO REL...	NIVEL RE...	VALOR N...	TIPO CUA...	NIVEL CU...	VALOR N...
1234567890	SANCHEZ ANDRA...	MASCULINO	INGENIERI...	B	9	ARQUITEC...	COMPAÑ...	EXCELEN...	0.93	TRABAJO...	CASI ALTO	0.90
1234567890	SANCHEZ ANDRA...	MASCULINO	INGENIERI...	B	9	ROBOTICA	COMPAÑ...	EXCELEN...	0.93	TRABAJO...	CASI ALTO	0.90

Figura 9. Consulta mediante aplicación después de transformación difusa

4. Discusión

Este trabajo ha permitido identificar atributos difusos de la base de datos de la tutoría académica, bajo un modelo relacional difuso, que permite extender el esquema tradicional de los sistemas de base de datos (objeto, atributo, valor) a la información imprecisa (en su representación) e incierta (en su consulta) con los conceptos de conjuntos difusos. El nuevo esquema, lo podemos representar como: (objeto, atributo difuso, valor), donde, los atributos difusos del modelo relacional

difuso, se encuentran clasificados en: Tipo I (valores precisos), Tipo II (valores imprecisos sobre referencial ordenado), y Tipo III (valores imprecisos sobre referencial no ordenado). Los valores, pueden ser: valores numéricos, escalares simples, etiquetas lingüísticas.

Los elementos de cada conjunto difuso, tienen un grado de pertenencia o grado de posibilidad, en SQL Server la herramienta Fuzzy Lookup permite calcular este grado, además permite traducir o transformar los datos difusos para poder procesarlos como datos clásicos, eliminando la complejidad de otros servidores FSQL que utilizan SQL extendido realizar consultas difusas.

La simulación de la imprecisión y la incertidumbre en una base de datos relacional difusa, es posible hacerlo en un sistema de información cartesiano (0 o 1), lo que implica la creación de múltiples tablas de conversión para simular el hecho que un valor se encuentra en el intervalo [0,1]

5. Conclusiones y Recomendaciones

Es necesario contar con un proceso matemático para analizar información difusa, debido a que, la mayoría de datos difusos son expresados en lenguaje natural a través de etiquetas lingüísticas, es conveniente asumir valores que representen grados de pertenencia en una escala de [0, 1] y en base a estos valores proceder con cualquier método de regresión lineal a encontrar los valores reales que sean bases en la transformación difusa

La definición del modelo de base de datos difusa, debe ser en base al DBMS que vaya a ser utilizado, ya que estos pueden basar sus transformaciones difusas en servidores FSQL o en herramientas CASE donde los resultados obtenidos tienen como fundamento umbrales de similitud, grados de confianza, distribuciones de posibilidades o únicamente grados de pertenencia de cada atributo difuso.

El sistema gestor utilizado permite analizar la información difusa en base a umbrales de similitud y grados de confianza, lo cual pone a disposición dos opciones para garantizar los resultados de las transformaciones difusas.

Al implementar el modelo de bases de datos difusos basados en puntuaciones de similitud y al definir un umbral de similitud del 50%, proveen un nivel de confianza del 0.9875, brindando la seguridad de que los atributos difusos que cumplan con esta condición puedan asumir los valores numéricos según la tendencia que tiene los atributos originales con respecto a los atributos de referencia que son la base de búsqueda en la herramienta Fuzzy Lookup.

SQL Server Integration Services – SSIS, es una herramienta muy útil que facilita el procesamiento de datos clásicos o difusos y permite integrarlos para ser sujetos de consulta o

procesamiento dentro del mismo ambiente de SQL, en SSIS es posible condensar dentro de un mismo flujo de procesos, tres actividades, automatizando en su mayoría el proceso de transformación difusa y delegando como proceso manual únicamente la ejecución de dicho flujo.

Bibliografía

- A. Silberschatz, H. K. (2002). *Fundamentos de Bases de Datos*. Madrid, España: Mcgraw-Hill/Interamericana De España.
- A. Urrutia, L. T. (2008). *FSQL and SQLF: Towards a standard in fuzzy database*.
- Barranco, C.D., Campaña, J.R., & Medina, J.M. (2008). *Towards a Fuzzy Object-Relational Database Model*. Recuperado el 2015, de Information Science Reference: <http://www.infosci-ref.com>
- D. Anderson, D. S. (2008). *Estadística para Administración y Economía*. México: Cengage Learning Editores.
- G. Bravo, G. R. (2012). Sistema de tutoría universitaria-SITUTOR. UTE.
- Network, M. D. (2016). Obtenido de Fuzzy Lookup Transformation: <https://msdn.microsoft.com/en-us/library/ms137786.aspx>
- Pérez, I., & León, B. (2007). *Lógica Difusa para Principiantes - Teoría y Práctica*. Caracas: Universidad Católica Andrés Bello.
- Pressaman. (2010). *Ingeniería de software*. Mexico.
- R.M. Cicília, G. L. (2011). A fuzzy model for processing and monitoring vital signs in ICU patients. *BioMedical Engineering OnLine*, 17.